2

TERMINOLOGY

complement
conditional probability
continuous variable
discrete variable
domain
event
expected value
failure
hypergeometric
independent
intersection
mutually exclusive
outcome
probability distribution
probability function
random
relative frequency
sample space
standard deviation
success
uniform distribution
union
variance

Prior learning

DISCRETE RANDOM VARIABLES

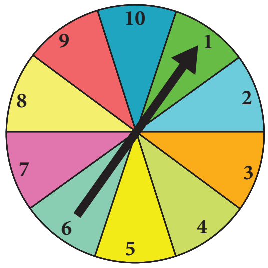
DISCRETE RANDOM VARIABLES

GENERAL DISCRETE RANDOM VARIABLES

- understand the concepts of a discrete random variable and its associated probability function, and their use in modelling count data (ACMMM136)
- use relative frequencies obtained from data to obtain point estimates of probabilities associated with a discrete random variable (ACMMM137)
- recognise uniform discrete random variables and use them to model random phenomena with equally likely outcomes (ACMMM138)
- examine simple examples of non-uniform discrete random variables (ACMMM139)
- recognise the mean or expected value of a discrete random variable as a measurement of centre, and evaluate it in simple cases (ACMMM140)
- recognise the variance and standard deviation of a discrete random variable as measures of spread, and evaluate them in simple cases (ACMMM141)
- use discrete random variables and associated probabilities to solve practical problems (ACMMM142) AC

2.01 DISCRETE RANDOM VARIABLES

In probability, an **experiment** has a number of **outcomes**. Each outcome is called a **sample point** and the **sample space** of an experiment is all possible outcomes. An **event** is a subset of the sample space.



Consider the spinner shown here. If event A is defined as 'an even number' and event B is defined as 'a multiple of 3', then:

$A = \{2, 4, 6, 8, 10\}$

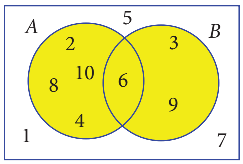
and $B = \{3, 6, 9\}$.

The **sample space** S is the set of individual possible occurrences, so in this case, $S = \{1, 2, 3, \ldots, 10\}$.

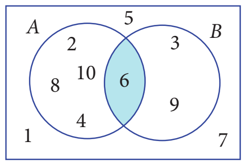
The **union** ($\cup$) of A and B is the combination of either event A or event B or both A and B occurring. So $A \cup B = \{2, 3, 4, 6, 8, 9, 10\}$.

The **intersection** ($\cap$) of A and B is the event that both A and B occur and includes the sample points that are common to A and B. So $A \cap B = \{6\}$.

Venn diagrams are used to visualise events, using circles in a rectangle to represent S.



$A \cup B$



$A \cap B$

The probability of an event is the likelihood or chance of it occurring, so $P(A) = \frac{5}{9}$ and $P(A \cup B) = \frac{7}{9}$

9780170250290

IMPORTANT

The **probability** of an event A in a sample space S is written as $P(A)$ and is a real number between 0 and 1. The probability of the sample space, $P(S) = 1$ and the probability of a union of disjoint events is the sum of their probabilities.

For a finite sample space whose elements have equal probabilities,

$$P(A) = \frac{\text{number of favourable outcomes}}{\text{total number of possible outcomes}} = \frac{n(A)}{n(S)}$$

If $P(A) = 1$, then event A is **certain** to occur.

If $P(A) = 0$, then event A is **impossible**.

The **complement** of event A is represented as A' or $\bar{A}$. A' means 'not A', so $P(A')$ is the probability that A will not occur.

$P(A) + P(A') = 1$

$P(A') = 1 - P(A)$

You can use Venn diagrams, tree diagrams, tables and grids to help calculate the probabilities of events. You can also use the following rules.

IMPORTANT

The **addition rule** of probability states that:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$P(A \cup B)$ is also written as $P(A \text{ or } B)$.

$P(A \cap B)$ is also written as $P(A \text{ and } B)$.

Mutually exclusive events cannot occur simultaneously, so $P(A \cap B) = 0$. For mutually exclusive events: $P(A \cup B) = P(A) + P(B)$

The **conditional probability** of event A given event B is written as $P(A|B)$ or 'the probability of A given B' and defined as $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

The **multiplication rule** for any events A and B is $P(A \cap B) = P(A|B) \times P(B)$.

Events are **independent** if $P(A|B) = P(A)$. The outcome of one does not affect the probability of the other, so $P(A \cap B) = P(A) \times P(B)$.

You should remember from Year 11 that conditional probability is often used when events occur one after another, but it can be used in unordered situations as well.

Example 1

Sam knows that if she wakes up when her phone alarm goes off, she has a 90% chance of getting to school on time. If she sleeps through the alarm, she only has a 40% chance of getting to school on time. Because she sometimes sleeps heavily, she only wakes up when the alarm rings 80% of the time. Calculate these probabilities.

iStockphoto/Jbryson

a She wakes up when the alarm rings and she gets to school late.
b She gets to school on time.
c She gets to school late.
d She wakes up when the alarm rings, given that she gets to school on time.

Solution

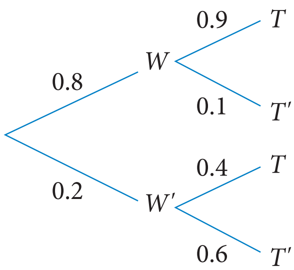
Define the events.

Let W = wakes up with alarm
Then sleeps through alarm = W'
Let T = arrives at school on time
Then arrives at school late = T'

Assign probabilities to each event.

$P(W) = 80\% = 0.8$
$P(W') = 1 - 0.8 = 0.2$
If she wakes up, $P(T|W) = 90\% = 0.9$
If she wakes up, $P(T'|W) = 1 - 0.9 = 0.1$
If she sleeps in, $P(T|W') = 40\% = 0.4$
If she sleeps in, $P(T'|W') = 1 - 0.4 = 0.6$

Draw a tree diagram and write the relevant probability on each branch.



a Write the required probability.

$P(\text{wakes up and gets to school late}) = P(W \cap T')$
$= P(T' \cap W)$

9780170250290

Use the multiplication rule. $= P(T'|W) \times P(W)$

$= 0.8 \times 0.1$

$= 0.08$

b Write the required probability. $P(\text{school on time}) = P(T)$

Use the addition rule. $= P(T \cap W) + P(T \cap W')$

Use the multiplication rule. $= P(T) \times P(T|W) + P(T) \times P(T|W')$

$= 0.8 \times 0.9 + 0.2 \times 0.4$

$= 0.72 + 0.08$

$= 0.8$

c Getting to school late is the complement of getting to school on time.

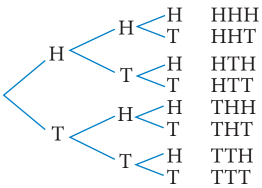
$$\begin{aligned} P(L) &= P(T') \\ &= 1 - P(T) \\ &= 1 - 0.8 \\ &= 0.2 \end{aligned}$$

d Use the rule for conditional probability.

$$\begin{aligned} P(W|T) &= \frac{P(W \cap T)}{P(T)} \\ &= \frac{0.72}{0.8} \\ &= 0.9 \end{aligned}$$

In Example **1**, the outcomes are occurrences in the real world. You can also have a sample space with numerical outcomes.

Consider a fair coin that is tossed three times and counting the number of tails in a row. You can use a tree diagram or list to work out the probabilities of each outcome.



iStockphoto/slobo

There are four different possibilities: 0, 1, 2 and 3 tails, but there are 8 equally likely elements in the sample space. The probabilities of tails in a row are shown below.

Number of tails in a row	Outcomes	Probability
0	HHH	$\frac{1}{8} = 0.125$
1	HHT, HTH, THH, THT	$\frac{1}{2} = 0.5$
2	HTT, TTH	$\frac{1}{4} = 0.25$
3	TTT	$\frac{1}{8} = 0.125$

In this case, the events are part of a sample space and have a numeric variable associated with them. The variable is the number of successive tails, T. You cannot tell what the value will be, so you say that it is **random**. It can only have the values 0, 1, 2 and 3, so it is a **discrete** variable, but you can work out the probability of each value.

IMPORTANT

A **random variable** is a variable with a numerical value determined by events from a sample space. Each value has an associated probability. A random variable is **discrete** if the possible values (the **domain**) are discrete. A **continuous random variable** has a continuous domain.

A random variable is usually denoted by a capital letter. The corresponding lower-case letter denotes specific values of the variable.

The **probability function** of a discrete random variable is the function formed by the values of the variable and their probabilities. For the random variable X, the probability function is shown as $P(X = x)$ or $p(x)$.

You can show a function as a list (set) of ordered pairs, table or rule. For the variable T above, $p(0) = 0.125$, $p(1) = 0.5$, $p(2) = 0.25$ and $p(3) = 0.125$.

WS

Classifying variables 1

Example 2

A pair of fair dice is tossed and the total of the uppermost faces is noted.

a If X = total of the uppermost faces, show that X is a discrete random variable.

b List the probability function.

Solution

a List the domain.

Domain of $X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

State how X is determined.

The values of X are determined by the sample space from tossing fair dice.

State the result.

X is both discrete and random.

b Construct a grid to determine the outcomes for this experiment.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

9780170250290

Calculate the probabilities of each value of the variable.
There are a total of 36 outcomes.

x	Number of outcomes	$p(x)$
2	1	$\frac{1}{36}$
3	2	$\frac{1}{18}$
4	3	$\frac{1}{12}$
5	4	$\frac{1}{9}$
6	5	$\frac{5}{36}$
7	6	$\frac{1}{6}$
8	5	$\frac{5}{36}$
9	4	$\frac{1}{9}$
10	3	$\frac{1}{12}$
11	2	$\frac{1}{18}$
12	1	$\frac{1}{36}$

List the probabilities.

$p(2)=\frac{1}{36}, p(3)=\frac{1}{18}, p(4)=\frac{1}{12}, p(5)=\frac{1}{9}, p(6)=\frac{5}{36},$
$p(7)=\frac{1}{6}, p(8)=\frac{5}{36}, p(9)=\frac{1}{9}, p(10)=\frac{1}{12}, p(11)=\frac{1}{18},$
$p(12)=\frac{1}{36}$

You could also show the probability function as a list of ordered pairs:

$(2, \frac{1}{36}), (3, \frac{1}{18}), (4, \frac{1}{12}), (5, \frac{1}{9}), (6, \frac{5}{36}), (7, \frac{1}{6}), (8, \frac{5}{36}), (9, \frac{1}{9}), (10, \frac{1}{12}), (11, \frac{1}{18}), (12, \frac{1}{36})$

or as the rule $p(x)=\begin{cases} \dfrac{x-1}{36} & \text{for} \quad 2 \le x \le 7 \\ \dfrac{13-x}{36} & \text{for} \quad 8 \le x \le 12 \end{cases}$.

EXERCISE 2.01 Discrete random variables

Concepts and techniques

WS Classifying variables 2

1 Classify each of the following variables as discrete or continuous.
   a The mass of a bus
   b The number of passengers in a bus
   c The number of bus stops on a route
   d The distance between bus stops
   e The price of a washing machine
   f The capacity of a washing machine
   g The energy star rating of a washing machine
   h The volume of water used by a washing machine in a wash cycle

2 List the domain for the variable in each of the following experiments.
   a A fair die is rolled. X = the number on the uppermost face.
   b A bag contains marbles numbered 1 to 10 and one marble is selected. X = the number on the selected marble.
   c A fair coin is tossed three times and X = the number of successive heads.
   d A card has a whole number between 1 and 10 (inclusive) written on it. A person is asked to guess the number. X = the number of guesses it takes to guess the number.
   e A point inside a circle with radius 2 units is chosen at random. X = the distance of the point from the centre of the circle.

3 For each experiment described in question **2**, classify the variable as either discrete or continuous and either random or non-random.

4 Which of the following random variables is not discrete?
   A The number of glass jars recycled by a family each fortnight.
   B The number of cars sold in a used car yard each month.
   C The number of runs scored by a batsman in each over of a match.
   D The height of a child measured each week from ages 5 to 10.
   E The price in cents of a kilogram of bananas at a supermarket each week for a year.

5 A driver must pass through 7 sets of traffic lights on the way to work each day in peak hour traffic. If N = the number of red lights the driver stops at, which of the following best describes N?
   A continuous variable
   B random variable
   C discrete variable
   D discrete continuous variable
   E random continuous variable

6 For the situation described in question **5**, which of the following represents the domain of the variable?
   A $\{7\}$
   B $\{0, 1, 2, \ldots, 7\}$
   C $0 \le n \le 7$
   D $\{1, 2, 3, 4, 5, 6, 7\}$
   E $\{1, 2, 3, \ldots, 6\}$

7 A bag contains 2 black marbles and 3 white marbles of identical size. If two marbles are drawn without replacement, what is the probability that they are different colours?
   A $\frac{6}{25}$
   B $\frac{3}{10}$
   C $\frac{2}{5}$
   D $\frac{12}{25}$
   E $\frac{3}{5}$

9780170250290

8 If $P(R) = 0.2$, $P(T) = 0.6$ and $P(T|R) = 0.5$, then $P(R \cup T)$ is equal to:

A 0.1 B 0.2 C 0.3 D 0.7 E 0.8

9 If $P(M) = 0.4$, $P(Q) = 0.5$ and $P(M \cup Q) = 0.7$, which one of the following is not true?

A M and Q are independent
B $P(M \cap Q) = 0.2$
C $P(M \mid Q) = 0.4$
D M and Q are mutually exclusive
E $P(Q \mid M) = 0.5$

10 Example 1 Jamal's car starts 80% of the time. He knows that if the car starts, he has a 90% chance of getting to work on time. If the car doesn't start, his chance of getting to work on time is only 40%. What is the probability that Jamal arrives to work late?

A 0.12 B 0.2 C 0.7 D 0.72 E 0.8

11 A factory that assembles computer motherboards has three assembly lines – A, B and C. It is known that 40% of the motherboards are assembled on line A and 30% are assembled on each of lines B and C. All motherboards are tested when they come off the line and defective ones are discarded. If a motherboard is assembled on line A, the probability of it being defective is 0.01. If it comes from line B, the probability of a defective motherboard is 0.02, while the probability is 0.03 if it comes from line C.

a If a newly assembled motherboard is randomly selected for testing, what is the probability that it is not defective?

b Given that a randomly selected motherboard is defective, what is the probability that it came from line B?

12 Example 2 A pair of tetrahedral dice with faces numbered 1 through 4 is tossed and the sum of the numbers on the faces on which the dice rest is calculated.
If X = total of the faces, list the ordered pairs that make up the probability function $P(X)$.

Shutterstock.com/Joyce Mar

Reasoning and communication

13 A fair coin is tossed four times. If H = the number of successive heads, calculate:

a $P(H = 3)$ b $P(H > 3)$ c $P(H < 3)$
d $P(H = 5)$ e $P(H \geq 2)$
f List the domain of M and the probability function.

14 A pair of cubic dice are thrown. If M = the minimum of the two numbers that occur, list the domain of M and the probability function $P(M = m)$.

15 A fair coin is tossed four times and T = the number of tails.

a List the domain of T and the ordered pairs that make up the probability function.
b Calculate $P(T > 2)$.

2.02 DISCRETE PROBABILITY DISTRIBUTIONS

It is often useful to display the values of a discrete random variable with their probabilities.

You usually draw the graph of a probability distribution like a histogram, without spaces between the columns, to emphasise the connection with statistical graphs. They are called probability histograms.

You can use probability properties to get the general properties of discrete probability distributions.

IMPORTANT

A **discrete probability function (discrete probability distribution)** shows the probability of a discrete random variable as a list, table or graph.

A discrete probability function $P(X = x)$ has the following properties.

- All the values are between 0 and 1: $0 \le P(X = x) \le 1$ for all values of x
- The sum of all the probabilities is 1. This is written as $\Sigma P(X = x) = 1$.
- $p(x)$ is a function, so every value of x has only one value of $p(x)$.

The Greek letter Σ is widely used in Mathematics to mean 'the sum of'. If there are no subscripts, then it means 'the sum of all values', as in the case above.

On the other hand, $\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$ and $\sum_{n=4}^{6} n^3 = 4^3 + 5^3 + 6^3$.

Provided that the values of X are discrete, the properties are sufficient to show that a function $p(x)$ is a discrete probability distribution. Of course, different values of x can have the same probability.

Example 3

A bag contains an equal number of identically shaped black and white discs. A disc is randomly selected, its colour noted and then it is returned to the bag. This is done three times in all.

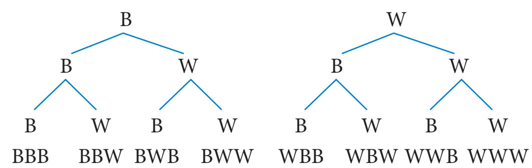
a Construct a probability distribution for the variable X, the number of black discs drawn.

b Confirm that it is a probability distribution function.

c Draw a graph of the probability distribution.

Solution

a Draw a tree diagram to calculate the outcomes for this experiment.



0, 1, 2 or 3 black discs can be drawn. List the outcomes for each event.

Number of black discs (x)	Outcomes
0	WWW
1	BWW, WBW, WWB
2	BBW, BWB, WBB
3	BBB

9780170250290

Calculate the probabilities for each value of the variable.

x	$p(x)$
1	$\frac{1}{8}$
2	$\frac{3}{8}$
3	$\frac{3}{8}$
5	$\frac{1}{8}$

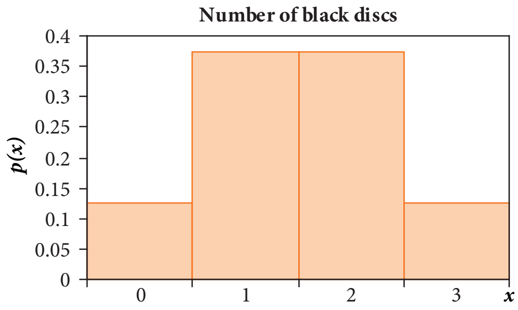
b Calculate the sum of the probabilities.

$\sum p(x) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$

State the result.

Since $0 \le P(X = x) \le 1$ for all values of x, $p(x)$ is a function and the sum of the probabilities is 1, it is a probability distribution function.

c Draw the graph of the probability distribution.



You can use the properties to determine if a function could be a probability function.

Example 4

A function is given as $p(x) = \frac{x}{15}$ for $x = 1, 2, 3, 4, 5$.

a Construct a table of values for the function.

b Determine if the function is a discrete probability distribution.

Solution

a Calculate the values.

x	$p(x)$
1	$\frac{1}{15}$
2	$\frac{2}{15}$
3	$\frac{3}{15}$
4	$\frac{4}{15}$
5	$\frac{5}{15}$

b Check the values.	$0 \le p(x) \le 1$ for all values of x.
Calculate the total.	$\sum P(X=x) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} + \frac{4}{15} + \frac{5}{15}$ $= 1$
Check that it is a function.	Each domain value has one value of $p(x)$.
State the conclusion.	The conditions are satisfied, so $p(x)$ is a discrete probability distribution.

Example 5

A discrete probability function is defined by the following table.

a Find the value of p.

b Hence calculate $P(X > 2)$.

x	0	1	2	3	4
$P(X=x)$	0.2	0.3	$2p$	p	0.05

Solution

a The sum of the probabilities must be 1.	$0.2 + 0.3 + 2p + p + 0.05 = 1$
Simplify.	$3p + 0.55 = 1$
Solve for p.	$p = 0.15$
b Write as a sum.	$P(X > 2) = P(X=3) + P(X=4)$
Use the value of p.	$= 0.15 + 0.05$
Write the answer.	$P(X > 2) = 0.2$

EXERCISE 2.02 Discrete probability distributions

Concepts and techniques

1 Example 3 A pair of dice is rolled and you win \$6 if the sum is greater than 10. If you get a double you win \$2 and if you get a double with a sum greater than 10 you win \$8. Otherwise, you win nothing. If X = amount won, which of the following is the probability distribution for X?

A

x	0	2	6	8
$P(X=x)$	$\frac{7}{9}$	$\frac{5}{36}$	$\frac{1}{18}$	$\frac{1}{36}$

B

x	0	2	6	8
$P(X=x)$	$\frac{3}{4}$	$\frac{1}{6}$	$\frac{1}{18}$	$\frac{1}{36}$

C

x	0	2	6	8
$P(X=x)$	$\frac{13}{18}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{36}$

D

x	0	2	6	8
$P(X=x)$	$\frac{27}{36}$	$\frac{5}{36}$	$\frac{1}{18}$	$\frac{1}{36}$

E

x	0	2	6	8
$P(X=x)$	$\frac{7}{12}$	$\frac{5}{36}$	$\frac{1}{12}$	$\frac{1}{36}$

9780170250290

2 Example 4 A function is defined as $p(x) = \dfrac{(x-2)^3}{35}$ for $x = 1, 2, 3, 4$. Choose the best of the following options.

A It is a probability function because the sum of the values is 1.
B It is not a probability function because $p(2) = 0$.
C It is a probability function because it is a function.
D It is not a probability function because $p(1) < 0$.
E Both **A** and **C** are true.

3 Example 5 If the table below represents a discrete probability distribution, calculate the value of m.

x	2	5	8	12
$P(X = x)$	$2m$	$5m$	$6m$	$3m$

A 0.01 B 0.0625 C 0.125 D 0.16 E 0.2

4 For the discrete probability distribution shown below, $P(x \le 8)$ is equal to:

x	3	5	8	9	11
$P(X = x)$	0.12	0.25	0.33	0.21	0.09

A 0.09 B 0.3 C 0.37 D 0.63 E 0.7

5 A fair coin is tossed three times. If X = number of tails in three tosses, then $p(x) = 2$ is equal to:

A $\frac{1}{8}$ B $p(x) = 1$ C $\frac{1}{4}$ D $p(x) = 3$ E $\frac{1}{2}$

6 For the discrete probability distribution shown below, calculate the value of n.

x	0	1	2	3	4
$P(X = x)$	0.15	0.25	n	0.35	0.05

A 0.15 B 0.2 C 0.25 D 0.35 E 0.4

7 State whether or not each of the following could represent a probability distribution.

a

x	$p(x)$
1	0.3
2	0.2
3	0.4

b

m	$p(m)$
−3	0.1
−1	0.2
1	0.3
3	0.4

c

t	$p(t)$
0	0.4
2	−0.1
6	0.2
8	0.5

8 Example 4 A probability distribution is defined by: $p(x) = \dfrac{x}{10}$ and $x = 1, 2, 3, 4$.

a Construct the probability distribution.
b Draw a graph of the distribution.
c Verify that it is a probability distribution.

9 Two probability distributions are defined by: $p_1 = \dfrac{x^2}{30}$ and $p_2 = \dfrac{5x^3}{12} - \dfrac{x^4}{24} - \dfrac{3}{4} - \dfrac{35x^2}{24} + \dfrac{25x}{12}$ for $x = 1, 2, 3, 4$.

a Construct the probability distributions.
b Verify that each is a probability distribution.
c Draw graphs of the distributions.

Reasoning and communication

10 A pair of tetrahedral dice with faces numbered from 1 to 4 are rolled and the sum (F) of the numbers on the faces on which the dice rest is calculated. Construct the probability distribution for F.

11 A pair of six-sided dice is rolled and the greater of the numbers on the upper faces, G, is noted.
- a Construct the probability function for G.
- b Draw a graph of the distribution.

12 A pair of six-sided dice is rolled and the sum of the numbers on the upper faces, S, is noted.
- a Construct the probability distribution for S.
- b Draw a graph of the distribution.
- c Verify that it is a probability distribution.

13 A coin is weighted so that $P(H) = \frac{2}{3}$ and $P(T) = \frac{1}{3}$, and it is tossed three times. Let X be the random variable representing the largest number of successive heads that occur.
- a Construct the probability distribution for X.
- b Draw a graph of the distribution.

14 A player rolls a fair six-sided die. If a prime number occurs, the player wins that number of dollars. If a non-prime number occurs, the player loses that number of dollars. If X is the number of dollars the player stands to win:
- a construct the probability distribution for X
- b draw a graph of the distribution.

15 A family consists of four children. Assuming that $P(\text{girl}) = \frac{1}{2}$, draw up a probability distribution for X, the number of girls in the family.

16 A game consists of rolling three six-sided dice. If all three dice show the same number, you win \$20. If two numbers are the same you win \$5, but if all three are different you lose \$5. If X is the random variable representing the amount you win, find the probability distribution of X.

2.03 ESTIMATING PROBABILITIES

You can use data to estimate probabilities. You need to count the number of times an event occurs in the data.

IMPORTANT

The **relative frequency** of an event is an estimate of its probability. It is given by

$$\text{Relative frequency} = \frac{\text{frequency of event}}{\text{total number of observations}} = \frac{\text{frequency of event}}{\text{total frequency}}.$$

The observations are usually called **trials**.

9780170250290

Example 6

1344 Australians aged 15–24 died in a 3-year period. 320 of the deaths were males killed in transport accidents. Over the same period, 389 of the females died. 1 356 900 males and 1 347 400 females were included in the study.

a Estimate the probability of an Australian man aged 15–24 dying in the 3-year period.

b What is the probability that the death of an Australian male aged 15–24 is due to a transport accident?

Solution

a Use the relative frequency. $\quad \text{Probability} = \dfrac{1344 - 389}{1\,356\,900}$

Simplify. $\quad \approx 0.0007$

Write the answer. $\quad$ The probability is about 0.0007 or 0.07%.

b Use the relative frequency. $\quad \text{Probability} = \dfrac{320}{1344}$

Simplify. $\quad \approx 0.238$

Write the answer. $\quad$ The probability is about 0.238 or 23.8%.

EXERCISE 2.03 Estimating probabilities

Concepts and techniques

1 Example 6 From the data in Example **6**, estimate the probability that the death of an Australian aged 15–24 is that of a female.

2 In another 3-year period, 286 of the deaths of 1309 Australians aged 15–24 were due to intentional self-harm, and 233 of these were males.
   - **a** What is the probability that the death of an Australian male in the age group 15–24 is due to self-harm?
   - **b** What is the probability that an Australian aged 15–24 dying from self-harm is a female?

3 From 40 people in a mall, 16 were observed to be wearing joggers. Estimate the probability that a person will be wearing joggers.

4 In 2005, 61 400 of the 85 200 babies born in Australia came from married women. Of those who were not married, 2300 chose not to name the father on the birth certificate. Use these figures to estimate the following.
   - **a** The probability that an Australian baby will be born to an unmarried mother.
   - **b** The probability that an Australian baby born to an unmarried mother will not have the father shown on the birth certificate.

5 In 2011, 198 300 babies were born to married women in Australia and of the 94 100 born to unmarried women, 10 100 did not name the father. Use these figures to estimate the following.
   - **a** The probability that an Australian baby will be born to an unmarried mother.
   - **b** The probability that an Australian baby born to an unmarried mother will not have the father shown on the birth certificate.

2.04 UNIFORM DISCRETE PROBABILITY DISTRIBUTIONS

In the previous sections the discrete probability functions were not constant. Cases where the probability function is constant occur quite frequently.

Example 7

A normal six-sided die is rolled and the uppermost number is noted.

a Draw a graph of the probability distribution.

b Calculate the probability that the number is greater than 4.

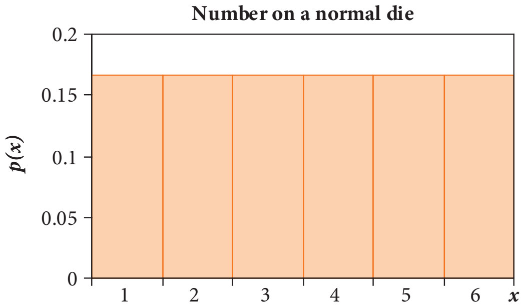
123RF/Aleksandar Kosev

Solution

a Construct the probability distribution.

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Draw the graph of the distribution.



b Write the probability.

$P(X > 4) = P(x = 5) + P(x = 6)$

Substitute values.

$= \frac{1}{6} + \frac{1}{6}$

Write the answer.

$= \frac{1}{3}$

The probability distribution in Example **7** is called a **uniform distribution**. The graph is a rectangle. The characteristics are as follows.

9780170250290

IMPORTANT

A **uniform discrete probability distribution** has equally likely outcomes.

For a domain with n elements, the probability function is $P(X = x) = \frac{1}{n}$.

If the domain is the integers from a to b inclusive, then $n = b - a + 1$, so $P(X = x) = \frac{1}{b-a+1}$.

It follows that, for this domain, $P(c \le x \le d) = \frac{d-c+1}{b-a+1}$, where $a \le c \le d \le b$.

A uniform distribution is often called a rectangular distribution.

If the clubs in a normal pack of cards are numbered from 11 to 23, the diamonds from 24 to 36, the hearts from 37 to 49 and the spades from 50 to 62, the domain is $\{11, 12, 13, \ldots, 62\}$ and $p(x) = \frac{1}{52}$.

Example 8

A bag contains identical discs numbered 10 to 19. A disc is drawn at random and returned to the bag. If R = the number on the disc drawn, calculate:

a $P(r = 12)$ **b** $P(15 \le r \le 17)$ **c** $P(x = 7)$

Solution

a There are 10 elements. $P(r=12)=\frac{1}{10}$

b $\{15, 16, 17\}$ has 3 out of 10 elements. $P(15 \le r \le 77)=\frac{3}{10}$

c 7 is not in the domain of $p(x)$. $P(x = 7)$ is not defined.

Some people prefer to say that the answer to part **c** of Example **8** is $P(x = 7) = 0$. You can use a slightly different definition of a discrete uniform probability distribution to allow for cases like this.

IMPORTANT

A **uniform discrete probability distribution** can be defined as a distribution that has n outcomes with equal probabilities, where $p(x) = \frac{1}{n}$ for $x \in D$ and $p(x) = 0$ for $x \notin D$.

Choosing a **random number** from 10 to 50 inclusive is a use of the discrete uniform probability distribution on the integers from 10 to 50.

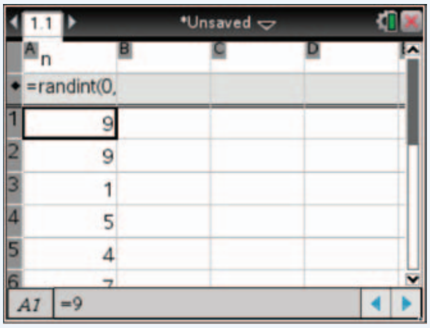
INVESTIGATION Random numbers

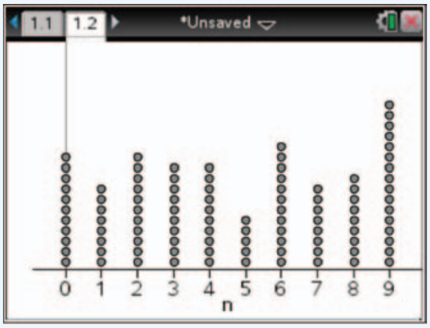
How random is your calculator?

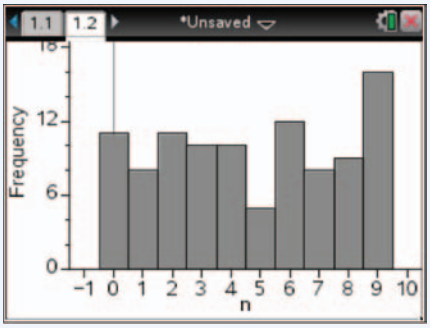
Many calculators have the facility to generate random numbers. Use your calculator to produce 100 random integers from 0 to 9 and plot a graph.

TI-Nspire CAS

Use the Lists & Spreadsheet page ( ).
Name the first list n and type
= randint(0, 9, 100) into the second row.
This generates 100 random integers from 0 to 9 in the A column.
Then press ctrl doc▾ and insert a Data & Statistics page ( ).
Move the cursor to the bottom and click on 'add variable' or press menu, 2: Plot Properties, 5: Add X Variable and choose n to see a graph.
Then press menu, 1: Plot Type and 3: Histogram to change the display to a histogram.

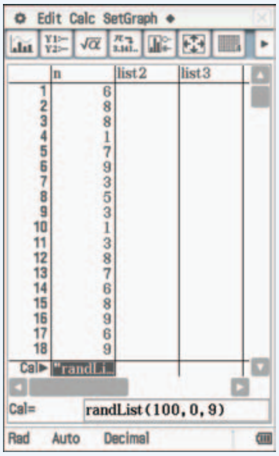




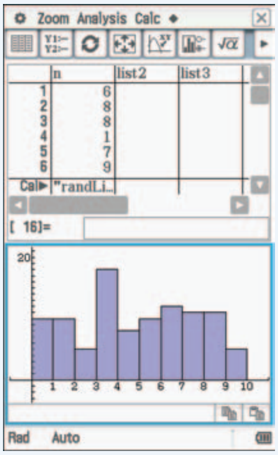


ClassPad

Use the Statistics menu (Statistics).
Tap the list1 cell and type n.
Tap the cell to the right of Cal▶, and then tap the cell to the right of Cal =. Enter randList(100, 0, 9) in this cell.
Tap SetGraph and Setting and select the Type Histogram. Choose main\n for the XList.
Freq should be set to 1. Tap Set.



Tap SetGraph and Stat Window Auto and set to off.
Then tap [histogram icon] to see the histogram. Set HStart to 0 and HStep to 1.
Tap [window icon] to set the window to appropriate scales.



Combine your results with those of other members of your class and comment on the probability distribution for the combined data.

EXERCISE 2.04 Uniform discrete probability distributions

Concepts and techniques

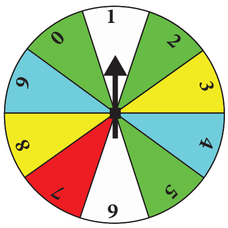
1 Example 7 A twelve-sided die is rolled and the uppermost number is noted. The probability of getting an 8 is:

A $\frac{1}{12}$ B $\frac{1}{8}$ C $\frac{1}{6}$ D $\frac{1}{4}$ E $\frac{8}{12}$

2 Example 8 A discrete random variable, X, has a uniform probability distribution. The domain of the variable is {4, 5, 6, 7, 8}. What is the probability that $x = 7$?

A $\frac{1}{8}$ B $\frac{1}{5}$ C $\frac{1}{7}$ D $\frac{4}{5}$ E 1

3 For each of the following situations describe the domain of the variable and state whether or not the variable would have a uniform probability distribution.



a A tetrahedral die is rolled and N = the number on the face on which the die rests.
b A spinner consists of 10 equal sectors, each numbered with a different digit. The spinner is spun and X = the number of the sector on which the arrow of the spinner stops.
c A circular disc consists of 10 equal sectors, each numbered with a different digit. People are asked to place their finger on the sector with their favourite digit. S = the number of the sector touched by each person.
d A bag contains 20 identical discs numbered 1 through 20. A disc is drawn at random, not replaced and then another disc is drawn. This process is repeated 10 times. D = the number on the disc.
e A fair die is rolled and M = the number on the uppermost face.
f Two fair dice are rolled and T = the total of their uppermost faces.

4 A bag contains 10 balls of identical size. The balls are numbered 1 through 10. A ball is randomly drawn from the bag. If N = the number on the ball, construct a probability distribution histogram and comment on its shape.

5 Eight identical discs, each with one of the numbers 1, 2, 3, 5, 7, 8, 10 and 11 are in a bag. A disc is drawn and then returned to the bag. If N is the number on the disc:

a construct a probability distribution histogram for N.

Use this histogram to calculate:

b $P(n \leq 4)$ c $P(n \neq 8)$ d $P(n \text{ is even})$

e $P(n \text{ is not even})$ f $P(2 \leq n \leq 8)$

6 A ten-sided die is rolled and X = the number on the uppermost face.

a Draw a graph of the probability distribution for X.

Use the graph of the probability distribution for X to calculate:

b $P(x \geq 2)$ c $P(x < 4)$ d $P(2 \leq x \leq 5)$

7 A spinner consists of 30 equal sectors numbered 1 through 30. The spinner is spun and S = the number of the sector on which the arrow of the spinner stops. Calculate the following values.

a $P(s = 17)$ b $P(s \neq 17)$ c $P(5 \leq s \leq 22)$

d $P(s > 17)$ e $P(s \leq 17)$

Reasoning and communication

8 A company sells goods online and customers ring a 1800 number to place an order. The time taken for the call centre to completely process the order is measured by a digital clock and recorded in minutes and seconds. The minimum unit of measurement is 1 second. It is known that the time taken to process and order can take as little as 150 seconds and as long as 12 minutes and that the times are randomly distributed over this range. If T = time taken to complete an order, calculate:

a $P(t > 8 \text{ min})$ b $P(t < 5 \text{ min})$

c $P(5 \text{ min} \leq t \leq 9 \text{ min})$ d $P(7 \text{ min} \leq t \leq 15 \text{ min})$

2.05 THE HYPERGEOMETRIC DISTRIBUTION

In this section, you will look at a simple non-uniform discrete probability distribution that frequently arises in quality control.

IMPORTANT

In a **hypergeometric experiment**, a sample of a particular number of items is taken from a finite population without replacement. The population is divided into two types, classified as success and failure, bad and good, or 1 and 0, with a fixed number of successes. The result of the experiment is the number of successes in the sample.

The sample size is usually written as n, the population size as N and the number of successes in the population is often written as k or M.

Example 9

People with O– blood are known as universal donors because their blood can be used in transfusions for people with any blood type without side effects. A researcher goes to a small school with 250 students. It is known from school records that 15 of the students have blood type O–. The researcher randomly selects a sample of 30 students and finds that 5 have O– blood.

a Is this an example of a hypergeometric experiment?

b List the values of the random variable X, the number of students who have O– blood type.

Solution

a	Is the population finite?	There is a finite population of 250.
	How is the sample taken?	A sample of 30 is taken without replacement.
	Are there a fixed number of successes?	The population is divided into two groups with 15 successes.
	All the conditions are met.	It is a hypergeometric experiment with $n = 30$, $N = 250$ and $k = 15$.
b	There are only 15 students with blood type O–.	$X = 0, 1, 2, 3, \ldots, 15$

As shown in Example **9**, there is a discrete random variable associated with hypergeometric experiments.

IMPORTANT

A **hypergeometric distribution** is the probability distribution of a hypergeometric experiment. The random variable is the number of successes. For a hypergeometric distribution with a sample size of n from a population of N with k successes, X has a domain from 0 to the minimum of n and k.

You can calculate the probabilities $P(X = x)$ in a hypergeometric probability distribution using the methods you learnt last year.

There are $\binom{k}{x}$ $({}^kC_x)$ ways of choosing x successes from the k successes in the population.

There are $\binom{N-k}{n-x}$ ways of choosing $n - x$ failures from the $N - k$ failures in the population.

Thus there are $\binom{k}{x}\binom{N-k}{n-x}$ ways of choosing a sample of n with k successes.

There are $\binom{N}{n}$ ways of choosing a sample of n from the population of N.

This gives the probability below.

IMPORTANT

For the hypergeometric random variable X, $P(X = x) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$, where $x \le n$ and $x \le k$.

Example 10

A box contains 50 electrical safety switches, of which 20% are known to be faulty. If a sample of 8 switches is selected at random and tested, calculate the probability that:

a two will be defective

b at least two will be defective.

Solution

a Identify the variable.

Let X = number of faulty switches

Identify the parameters.

$N = 50,\ n = 8,\ k = 0.2 \times 50 = 10$

Write the formula.

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Substitute the parameters and $x = 2$.

$$P(X = 2) = \frac{\binom{10}{2}\binom{50-10}{8-2}}{\binom{50}{8}}$$

Evaluate.

$$= \frac{\binom{10}{2}\binom{40}{6}}{\binom{50}{8}}$$

$$= 0.321\ 724\ldots$$

State the result.

The probability of two faulty switches is about 0.3217.

b Find the complement.

$P(X < 2) = P(X = 0) + P(X = 1)$

Use the formula.

$$= \frac{\binom{10}{0}\binom{40}{8}}{\binom{50}{8}} + \frac{\binom{10}{1}\binom{40}{7}}{\binom{50}{8}}$$

Calculate the value.

$$= 0.490\ 502\ldots$$

Find the complement.

$$P(X \geq 2) = 1 - 0.490\ 502\ldots$$
$$= 0.509\ 497\ldots$$

State the result.

The probability of at least two faulty switches is about 0.5095

One area where hypergeometric probability distributions are used is known as acceptance testing. This is a process in manufacturing where a sample of the output of a full production batch is tested to determine if the entire batch is acceptable.

9780170250290

Example 11

A computer manufacturer orders 500 printer circuit boards from a supplier. To determine whether or not to accept the order, the manufacturer randomly selects 15 circuit boards and tests them. The manufacturer will only accept the full order if there are no defective circuit boards in those that have been randomly selected for testing. Find each of the following probabilities.

a The probability of no faulty boards if 5% of the circuit boards are defective.
b The probability of no faulty boards if 10% of the circuit boards are defective.

Solution

a Identify a success.

Success = a defective circuit board

Identify the parameters.

$N = 500$, $n = 15$, $k = 0.05 \times 500 = 25$

Find $P(X = 0)$.

$$P(X=0) = \frac{\binom{25}{0}\binom{500-25}{15-0}}{\binom{500}{15}}.$$

Evaluate.

$$= 0.458\,096\ldots$$

State the result.

When 5% of all boards are defective, the probability that none of the sample boards is defective is about 46%.

b Identify the parameters.

$N = 500$, $n = 15$, $k = 0.1 \times 500 = 50$

We want to know $P(X = 0)$.

$$P(X=0) = \frac{\binom{50}{0}\binom{500-50}{15-0}}{\binom{500}{15}}$$

Evaluate.

$$= \frac{\binom{50}{0}\binom{450}{15}}{\binom{500}{15}}$$

$$= 0.201\,045\ldots$$

State the result.

When 10% of all boards are defective, the probability that none of the sample boards is defective is about 20%.

In Example **11**, the manufacturer would obviously prefer to know the probability of an acceptable number of faulty circuit boards if the number in the sample was zero, but this would require much more extensive calculations.

EXERCISE 2.05 The hypergeometric distribution

Concepts and techniques

WS Hypergeometric probability experiments

1 A hypergeometric probability experiment with a population of 50 has a random variable, X. If the sample size is 5 and it is known that there are 12 successes in the population, what is the maximum value of x?

A 5 B 7 C 12 D 38 E 50

2 Example 9 Which of the following is not a characteristic of a hypergeometric probability experiment?

A The population to be sampled must be finite.
B For each trial of the experiment, there are only two possible outcomes.
C A finite sample is randomly selected from the population.
D The probability of success in each trial is constant.
E The exact number of successes in the population is known.

3 A hypergeometric probability experiment with $N = 200$ has a random variable Y. If $n = 15$ and $k = 10$, what is the maximum value of y?

A 5 B 10 C 15 D 25 E 200

Use the hypergeometric probability experiment defined by $P(X = 3) = \dfrac{\binom{12}{3}\binom{38}{2}}{\binom{50}{5}}$ to answer questions **4** and **5**.

4 What is the sample size for the experiment?

A 2 B 5 C 12 D 38 E 50

5 What is the number of success in the population?

A 2 B 5 C 7 D 12 E 38

6 Example 9 For each of the following hypergeometric experiments, identify the random variable (X) and determine the values of N, n, k and the possible values of X.

a Four cards are drawn from a well shuffled standard deck of playing cards and the number of hearts selected is noted.
b An electrical appliance has 6 transistors. It is known that two of the transistors are faulty but it is not known which two. Three transistors are randomly selected and removed to test to see if they are faulty.
c It is known that 18% of a batch of 50 computer chips are defective. A computer manufacturer randomly selects a sample of 10 chips for testing before the batch is purchased to see if any are defective.
d A production run of 200 printed circuit boards contains 10% that are defective. 40 of the circuit boards are randomly selected to see if they are defective.
e A large office area has 95 workstations and 35 of these are each equipped with only a laptop computer. The remainder have only desktop computers. The laptops require a special login code in order to access the company's network. 12 staff are randomly assigned to a workstation. Each of the staff members is then asked if they require a special login code.

7 For the hypergeometric experiments described in question **6**, calculate the probability that:

a two of the four cards drawn are hearts
b one of the three transistors selected is faulty

9780170250290

c four of the computer chips tested are defective
d 10 of the printed circuit boards selected are defective
e three of the staff who are asked require a special login code.

8 A hypergeometric probability experiment is conducted with the given parameters below. Find the probability correct to 4 decimal places of obtaining x successes in each case.
a $N = 120, n = 20, k = 30, x = 7$
b $N = 50, n = 15, k = 10, x = 3$
c $N = 200, n = 50, k = 40, x = 9$
d $N = 30, n = 8, k = 5, x = 3$
e $N = 150, n = 25, k = 40, x = 11$

Reasoning and communication

9 Six cards are drawn from a standard deck of playing cards without replacement. What is the probability of drawing two spades?

10 Example 10 A batch of 100 circuit breakers contains 15% that are defective. If a sample of 19 circuit breakers is selected at random and tested, calculate the probability that:
a three will be defective
b at least three will be defective.

11 In a group of 200 people, 12 are known to suffer from colour blindness. A researcher randomly selects a sample of 20 people from the group and tests them.
a What is the probability that exactly three are colour blind?
b What is the probability that at least one is colour blind?

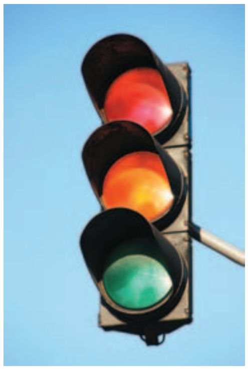
12 In the game of Lotto, a player picks 6 numbers from the numbers 1, 2, 3, …, 45 and marks them on a card. This is called an entry. Six balls numbered 1 to 45 are then randomly selected by the Lotto organisers. Prizes are won by players who select 3, 4, 5 or 6 numbers that match the ones that have been randomly selected. The jackpot (major prize) is won if the player has 6 matching numbers.
a What is the probability that a single entry will win the jackpot?
b What is the probability that a single entry will win a prize?

13 Example 11 A manufacturer received an order of 250 machined components. The components are only acceptable if they are within ±1 mm of the specified dimensions. It is known that 12 of the components are not within ±1 mm of the required dimensions and are therefore unusable. The order will be rejected if there are four or more unusable components. The manufacturer decides to randomly select 20 components for testing to see if they are within the specified dimensions.
a What is the probability that no unusable components are found?
b What is the probability that three unusable components are found?
c What is the probability that the order will be rejected?

14 In certain criminal trials, a jury is required to reach a unanimous verdict in order to convict an accused person. If a unanimous verdict is not reached, the jury is said to be 'hung' and the trial is abandoned. A particular jury consists of 12 people randomly selected from a pool of 50 potential jurors, of which 3 would never be willing to convict, regardless of the evidence presented at the trial. What is the probability that the trial will result in a hung jury, regardless of the evidence presented?

2.06 EXPECTED VALUE

Probability can be an important part of the decision-making process for a broad range of activities.



123RF/alinochka5

Example 12

Susan drives through 8 sets of uncoordinated traffic lights on her way to work. She needs to stop at the following numbers of red lights on 20 successive days on her way to work.

4	5	4	5	5	1	3	4	5	8	2	4	2
6	4	4	1	5	4	6						

a Estimate the probability distribution for R, the number of red lights obtained.

b Calculate the average number of red lights she needs to stop at on a trip.

Solution

a Make a frequency table and estimate the probabilities as relative frequencies.

Red lights, r	Frequency	$P(R = r)$
0	0	0.00
1	2	0.10
2	2	0.10
3	1	0.05
4	7	0.35
5	5	0.25
6	2	0.10
7	0	0.00
8	1	0.05
Total	**20**	**1**

b Work out the average number of red lights.

$$\text{Average} = \frac{4+5+4+5+5+1+\ldots}{20}$$

Calculate the answer.

$$= 4.1$$

State the result.

There is an average number of 4.1 red lights per trip.

WS Expected values

In Example **9**, the average number of red lights is the number of red lights Susan would (on average) expect to get on her way to work. This value can be worked out using the estimated probabilities. Each value of $p(r)$ above has been calculated by dividing by the total frequency, so $p(r) = \frac{f}{\sum f}$ for each value of r.

9780170250290

You already know that for the variable X, the **mean**, $\bar{x}$, is given by

$$\bar{x} = \frac{\sum fx}{\sum f}$$

You can rewrite this as $\bar{x} = \sum x \times \frac{f}{\sum f} = \sum x \cdot p(x)$

In the case of the traffic lights, this gives

$$\bar{x} = 0 \times 0.00 + 1 \times 0.10 + 2 \times 0.10 + 3 \times 0.05 + 4 \times 0.35 + 5 \times 0.25 + 6 \times 0.10 + 7 \times 0.00 + 8 \times 0.05$$

$$= 4.1.$$

You get the same result as adding individual scores or by using frequencies. Expected value is defined in this way for probability distributions in general, not just for those arising from experimental probabilities.

IMPORTANT

The **expected value**, μ or $E(X)$, of the probability distribution of the discrete random variable X is the mean value of X. For a discrete probability distribution, it can be calculated using the formula $\mu = E(X) = \sum xp(x)$.

Example 13

Find the expected value for the following distribution.

x	1	2	3	4	5
$P(X = x)$	0.2	0.1	0.3	0.1	0.3

Solution

Use the rule for expected value. $E(X) = \sum xp(x)$

Use the values from the distribution. $= 1 \times 0.2 + 2 \times 0.1 + 3 \times 0.3 + 4 \times 0.1 + 5 \times 0.3$

Evaluate and state the result. $= 3.2$

In Example **10**, the expected value for the distribution is not an actual value in the domain of the discrete random variable. This is because $E(X)$ is the theoretical mean.

A graphics calculator can be used to calculate expected value.

Consider the probability distribution shown here.

x	100	200	400	700	900
$P(X = x)$	0.2	0.1	0.3	0.1	0.3

You can use a graphics calculator to find the expected value of the distribution as follows.

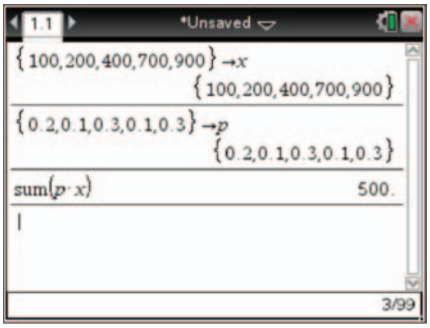
TI-Nspire CAS

Use the Calculator page.

Type in {100, 200, 400, 700, 900} and press ctrl var (sto→) x.

Type in {0.2, 0.1, 0.3, 0.1, 0.3} sto→ p.

Then add the products by typing in sum(p $\boxed{\times}$ x).

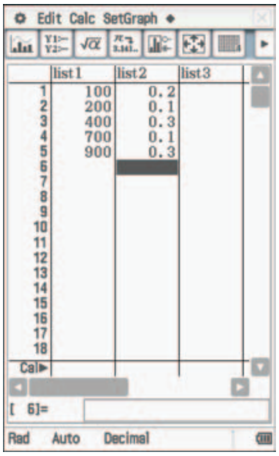


ClassPad

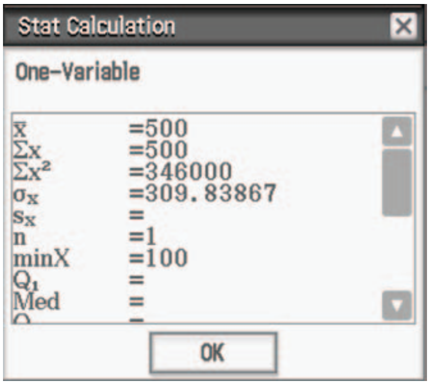
Use the Statistics menu and find the mean in the usual way.

Enter 100, 200, 400, 700, 900 in list 1 and the probabilities (instead of the frequencies) in list 2.

Tap Calc and choose One Variable. Set Xlist to list1 and Freq to list2.



Select the mean, $\bar{x} = 500$.



There are also other ways to use your CAS calculator to find expected values.

9780170250290

EXERCISE 2.06 Expected value

Concepts and techniques

Expected values using a calculator

Using expected values

1 Example 12 The results of a test out of 40 are shown below.

22 36 12 16 35 30 25 28 26 20 18 15 32
27 21 13 8 30 17 27

Estimate the expected result for the test.

A 11.45 B 21.8 C 22.9 D 23.4 E 24.1

2 Example 13 A discrete random variable X can take the values 0, 1 and 2 with probabilities 0.2, 0.5 and 0.3 respectively. What is the expected value of X?

A 0.8 B 0.9 C 1 D 1.1 E 1.2

3 The following is the estimated probability distribution for the number (N) of laptop computers in a household in a particular town. Estimate the expected value of N.

n	0	1	2	3	4
$p(n)$	0.2401	0.4116	0.2646	0.0756	0.0081

A 1.10 B 1.20 C 1.35 D 1.40 E 1.44

4 Calculate $E(X)$ for the following probability distribution.

x	0	5	10	15
$P(X = x)$	0.5	0.1	0.2	0.2

5 Find the expected value for each of the following distributions.

a

n	2	3	11
$p(n)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

b

n	−5	−4	1	2
$p(n)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

c

n	1	3	4	5
$p(n)$	0.4	0.1	0.2	0.3

6 A random variable has the following distribution.

x	0	1	2	5
$P(X = x)$	0.1	0.2	0.3	0.4

Calculate $E(X)$.

7 The results in a subject, out of 100, are shown below.

32 45 68 90 45 63 55 17 50 60 36 67 70
85 55 65 67 57 43 80

a Find the expected value of the variable F for the test results.

The test results are rescaled to a mark M out of 30.

b Find the expected value of the variable M.

8 CAS Calculate the expected value of the variable in each of the following distributions.

a

x	1200	1300	1400	1500	1600
$p(x)$	0.056	0.189	0.274	0.296	0.185

b

y	300	310	320	330	340	350	360
$p(y)$	0.268	0.333	0.162	0.14	0.062	0.023	0.012

9 A random variable has the following probability distribution.

x	1	2	3	4
$p(x)$	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{6}$

Calculate $E(X)$

Reasoning and communication

10 A coin is tossed three times. Let X = the largest number of successive heads.

a Calculate $E(X)$ if the coin is fair.

b Calculate $E(X)$ if the coin is weighted so that $P(\mathrm{H}) = \frac{2}{3}$.

11 Two dice are rolled and S, the equal or smaller of the two numbers on the dice, is noted.

a Draw up the probability distribution.

b Calculate the expected value.

12 Five identically-shaped balls numbered 1 to 5 are held in a bag. A ball is drawn and not replaced and then another ball is drawn. If S = the sum of the numbers drawn, calculate:

a the probability distribution for S

b $E(S)$

13 A game is played in which a pair of dice is rolled and a counter moved forward 1, 2 or 4 places according to the numbers shown on the dice. The counter is moved 1 place forward if the numbers differ by three or more, 2 places if they differ by one or two and 4 places if they are equal. How many places would you expect the counter is moved by for each roll of the dice?

14 A random variable X has the following probability distribution.

x	1	2	3	4	5
$P(X = x)$	$8k$	$5k$	$4k$	$2k$	k

a Calculate the value of k.

b Without calculating its value, explain why you expect $E(X)$ to be 3, greater than 3, or less than 3.

c Calculate $E(X)$.

15 A random variable W has the following probability distribution.

w	2	3	5	8	12
$P(W = w)$	$\frac{1}{8}$	$\frac{1}{3}$	$\frac{1}{4}$	x	y

If $E(X) = 5\frac{2}{3}$, find the values of x and y.

9780170250290

16 A fair coin is tossed until a head or 5 tails occur. If X is a random variable representing the number of tosses of the coin, find the expected value of X.

17 In a 6-cylinder car engine, 2 spark plugs are defective. Three spark plugs are removed at random and checked during a tune-up. If X is the number of defective spark plugs found, calculate the expected number of defective spark plugs found.

2.07 VARIANCE AND STANDARD DEVIATION

The expected value of a discrete random variable is a measure of its average value. In this section you will look at measuring the spread of a discrete random variable. Intuitively, the distance of a value from the mean measures how far apart that item is from the standard result, so the average distance from the mean would be a measure of spread. However, if you added the values of $x - \mu$, the ones above the mean would cancel out the ones below. Squaring the distance from the mean makes them all positive so the sum is not zero, and emphasises the values further from the mean. Hence the average of the squares of the distances from the mean gives a reasonable measure of spread. This is $E([X-\mu]^2)$. The square root of this value will give a value on the same scale as the original variable. $E([X-\mu]^2)$ can be converted to a more convenient form for calculation as follows.

$$\begin{aligned}
E\left([X-\mu]^2\right) &= \sum [X-\mu]^2 p(x) \\
&= \sum X^2 p(x) - \sum 2X\mu p(x) + \sum \mu^2 p(x) \\
&= E\left(X^2\right) - 2\mu \sum X p(x) + \mu^2 \sum p(x)
\end{aligned}$$

Since $\mu = \sum X p(x)$ and $\sum p(x) = 1$, then

$$\begin{aligned}
E\left(X^2\right) - 2\mu \sum X p(x) + \mu^2 \sum p(x) &= E\left(X^2\right) - 2\mu\mu + \mu^2 \\
&= E\left(X^2\right) - 2\mu^2 + \mu^2 \\
&= E\left(X^2\right) - \mu^2
\end{aligned}$$

Hence $E\left([X-\mu]^2\right) = E\left(X^2\right) - \mu^2$.

IMPORTANT

For the discrete random variable X with expected value $E(X) = \mu$, the **variance** is given by

$$\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E(X^2) - \mu^2
\end{aligned}$$

and the **standard deviation** is

$$SD(X) = \sigma = \sqrt{Var(X)}$$

Example 14

A die is rolled and $X =$ the square of the score. Calculate the mean, variance and standard deviation of X.

Solution

Show the probability distribution for X.

x	1	4	9	16	25	36
$P(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Calculate $E(X)$.

$$E(X) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + \cdots$$

Work out the answer.

$$\mu = 15\frac{1}{6}$$

Now calculate $E(X^2)$.

$$E(X^2) = 1 \times \frac{1}{6} + 16 \times \frac{1}{6} + 81 \times \frac{1}{6} + \cdots$$

Work out the answer.

$$= 379\,\frac{1}{6}$$

Write the rule for variance.

$$Var(X) = E(X^2) - \mu^2$$

Substitute in the values.

$$= \frac{2275}{6} - \left(\frac{91}{6}\right)^2$$

Work out the answer.

$$= \frac{5369}{36}$$

$$\approx 149.139$$

Calculate the standard deviation.

$$\sigma = \sqrt{\frac{5369}{36}}$$

$$\approx 12.212$$

State the result.

The mean is $15\,\frac{1}{6}$, the variance is about 149.1 and the standard deviation is about 12.2.

123RF/Paul Vasarhelyi

9780170250290

Example 15

The number of coins given in change to customers at a fast food restaurant are recorded below.

Coin	5c	10c	20c	50c	\$1	\$2
Number	143	151	215	52	141	98

a Estimate the probability distribution of X, the value of coins given in change.
b Estimate the expected value, variance and standard deviation of X.

Solution

a Use relative frequencies for the probability distribution.

x	0.05	0.1	0.2	0.5	1	2
$P(x)$	0.1788	0.1888	0.2688	0.0650	0.1763	0.1225

b Calculate $E(X)$.
$E(X) \approx 0.05 \times 0.1778 + 0.1 \times 0.1888 + \cdots$

Work out the answer.
$\mu \approx 0.5353$

Now calculate $E(X^2)$.
$E(X^2) \approx (0.05)^2 \times 0.1788 + (0.1)^2 \times 0.1888 + \cdots$

Work out the answer.
≈ 0.6956

Write the formula.
$Var(X) = E(X^2) - \mu^2$

Substitute in the values.
$= 0.6956 - (0.5353)^2$

Calculate the result.
≈ 0.4090

Find the standard deviation.
$SD(X) = \sqrt{Var(X)}$
≈ 0.6395

Write the answer.
The expected value is about 53 cents, the variance is about 0.41 and the standard deviation is about 64 cents.

In Example **15**, the values of the discrete random variable are somewhat disparate, rather than being consecutive integers.

WS Variance and standard deviation

Example 16

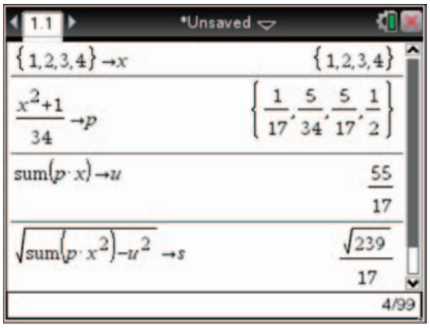
CAS A probability distribution for the discrete random variable, X, is given by $P(X = x) = \frac{x^2+1}{34}$ when $x = 1, 2, 3, 4$. Work out the following.

a The probability distribution.
b The expected value of X, correct to 3 decimal places.
c The standard deviation of X, correct to 3 decimal places.

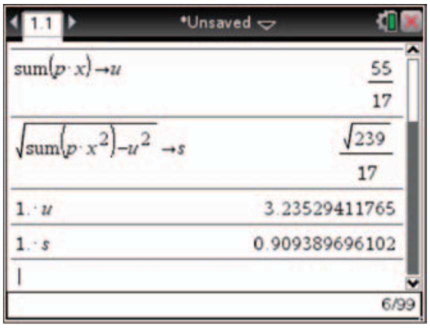
Solution

TI-Nspire CAS

a Use the Calculator page.
Store {1,2,3,4} in x.
Store the probabilities in p.



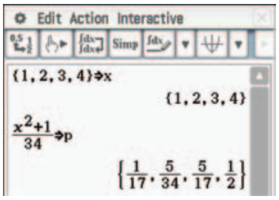
b Calculate the expected value using sum $(p \times x)$ and store in u.



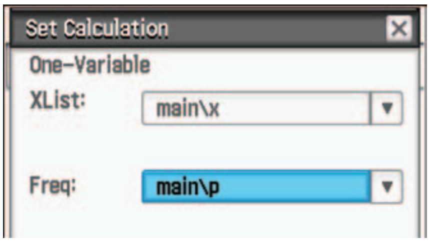
c Calculate the standard deviation using $\sqrt{\text{sum}(p \times x^2) - u^2}$ and store in s.
You switch to approximate calculation and multiply by 1 to approximate.

ClassPad

a Start with the Main menu. Set the calculator to Standard.
Store {1,2,3,4} in x.
Store the probabilities in p.

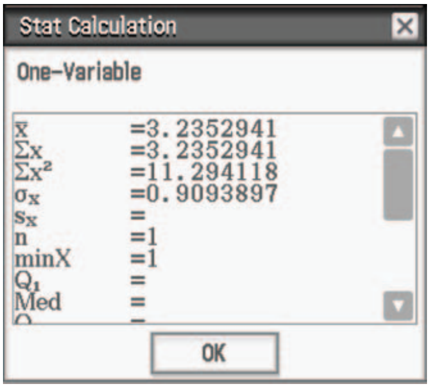


b Go to the Statistics menu.
and Rename list1 as x and list2 as p.



9780170250290

c The values automatically appear.
Tap Calc, One-Variable and make XList main\x and Freq main\p.
Read the mean ($\bar{x}$) and standard deviation (σ_x).
You can also use the method as shown in the TI-Nspire CAS.



Write the answers.

The probability distribution is $P(1) = \frac{1}{17}$, $P(2) = \frac{5}{34}$, $P(3) = \frac{5}{17}$, $P(4) = \frac{1}{2}$. The expected value is about 3.24 and the standard deviation is about 0.91.

The list processing shown in the last example is a powerful tool of CAS calculators.

EXERCISE 2.07 Variance and standard deviation

Concepts and techniques

1 Example 14, 15 The probability distribution for a random variable, Z, is shown below.

z	1	2	3	4
$p(z)$	0.1	0.2	0.3	0.4

The variance of Z is:

A 1.0 **B** 1.2 **C** 1.6 **D** 1.8 **E** 2.0

2 The probability distribution for a random variable, Y, is shown below.

y	100	200	300	400
$p(y)$	0.2	0.4	0.3	0.1

The standard deviation of Y is:

A 90 **B** 99 **C** 100 **D** 108 **E** 117

3 The probability distribution for a random variable, M, is shown below.

m	1	2	3	4
$p(m)$	a	$2a$	$3a$	$4a$

The standard deviation of M is:

A 1.0 **B** 1.2 **C** 1.6 **D** 1.8 **E** 2.0

4 The probability distribution of W is shown below.

w	0	2	4	6
$P(W=w)$	0.1	0.3	0.4	0.2

Find the variance and standard deviation of W.

5 For a discrete random variable X, $E(X) = 14$ and $E(X^2) = 280$. Find the variance of X.

6 For a discrete random variable Y, $E(Y) = 23$ and $E(Y^2) = 587$. Find the standard deviation of Y.

7 The values of a random variable, X, are recorded as 7, 6, 7, 3, 7, 6, 7, 8, 7, 6, 4, 7, 8, 7, 9, 8, 7, 9, 7, 8.

a Construct a probability distribution for X.

b Calculate $E(X)$.

c Calculate $Var(X)$.

8 Thirty shopkeepers in a coastal town were asked how many additional staff they intended to hire for the next holiday season. Their responses are shown below.

4, 5, 7, 9, 8, 2, 4, 2, 5, 3, 6, 7, 4, 4, 5, 6, 5, 5, 7, 8, 8, 0, 4, 6, 1, 9, 3, 3, 4, 6

If N = number of additional staff, calculate:

a the probability distribution for N

b $E(N)$

c $Var(N)$

d the standard deviation.

9 Calculate the mean (μ), variance and standard deviation (σ) for each of the following distributions.

a

n	−1	0	1	2	3
$p(N=n)$	0.1	0.3	0.1	0.2	0.3

b

r	11	12	13	14	15	16	17
$p(R=r)$	0.12	0.18	0.21	0.19	0.16	0.11	0.03

10 Example 16 CAS A probability distribution for the discrete random variable, Z, is given by

$P(Z=z) = \dfrac{z^2+2}{22}$ when $z = 0, 1, 2, 3$. Find the following.

a The probability distribution of Z.

b The expected value of Z, correct to 3 decimal places.

c The standard deviation of Z, correct to 3 decimal places.

11 CAS A discrete random variable has the following probability distribution.

y	2	4	6	8	10
$p(y)$	0.16	0.23	0.31	0.18	0.12

Calculate $Var(Y)$

9780170250290

12 CAS Calculate the mean (μ), variance and standard deviation (σ) of each of the following distributions.

a

x	1200	1300	1400	1500	1600
$p(x)$	0.056	0.189	0.274	0.296	0.185

b

y	300	310	320	330	340	350	360
$p(y)$	0.268	0.333	0.162	0.14	0.062	0.023	0.012

13 CAS The weights (to the nearest gram) of eggs laid by a number of hens are recorded as follows.

x	51	53	56	59	60	61	66	69
$p(X = x)$	0.05	0.08	0.1	0.23	0.25	0.16	0.09	0.04

where X = the weight of an egg.

Calculate:

a $E(X)$ b $Var(X)$ c σ

Reasoning and communication

14 A discrete random variable, W, has the following probability distribution.

w	1	3	k	6
$P(W = w)$	0.2	0.4	0.1	0.3

Find the value of k, a positive integer, if the variance is 3.41.

15 A discrete random variable, Y, has the following probability distribution.

y	1	k	7	11
$P(Y = y)$	0.1	0.3	0.4	0.2

Find the value of k, a positive integer, if the variance is 10.6.

16 A pair of dice is rolled and the sum, S, of the numbers on the uppermost faces is recorded. Calculate

a $E(S)$ b $Var(S)$ c σ

17 A fair six-sided die has a '2' on one face, '4' on two faces and '6' on the remaining three faces. The die is rolled twice and T = total of the two numbers rolled.

a Construct a probability distribution for T.

b Calculate the values of $E(T)$ and $Var(T)$.

18 Three cards are numbered 1, 2 and 3. The cards are placed face down, one is randomly selected, its number is noted and the card is returned face-down to the table. This procedure is then repeated. Let X = the sum of the two numbers and Y = the smaller of the two numbers.

a Construct a probability distribution for each variable.

b Calculate the values of $E(X)$ and $Var(X)$.

c Calculate the values of $E(Y)$ and $Var(Y)$.

2.08 APPLICATIONS OF DISCRETE RANDOM VARIABLES

You can use the expected value of a distribution to work out if a game of chance is fair. In a fair game of chance, your expected value would be 0. This is because you would expect to come out even in the long run.

Example 17

A dice game involves rolling a pair of dice and finding the total of the numbers on the upper faces. If the total is 11 or 12, you win $10. If the total is 2, 3 or 4, you win $4. For any other total, you lose $2. Is the game fair?

Solution

First calculate the probabilities for the totals using a grid.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(11 \text{ or } 12) = \frac{3}{36} = \frac{1}{12}$

$P(2, 3 \text{ or } 4) = \frac{6}{36} = \frac{1}{6}$

$P(\text{others}) = \frac{27}{36} = \frac{3}{4}$

Choose the random variable.

Let X be the amount won on a roll.

Write out the probability distribution.

x	10	4	−2
$P(x)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{3}{4}$

Work out the expected value.

$E(X) = 10 \times \frac{1}{12} + 4 \times \frac{1}{6} + (-2) \times \frac{3}{4} = 0$

Write the answer.

The game is fair because you win $0 (and lose $0) in the long run, so you 'break even'.

9780170250290

Casinos, poker machines and other gambling institutions have the payments structured unfairly so the expected value for the player is negative. The proportion they expect to win is usually expressed as a percentage and is called the **house percentage**.

Example 18

A game consists of a wheel divided into twelve equal sectors numbered 1 through 12. You may place a \$2 bet on any number. If your number comes up, you get \$18 back. That is, you win \$16 and get your \$2 back.

a What is the house percentage?

b How much should you win for the game to be fair?

Solution

a Choose the random variable.

Let X be the amount won on a bet.

Write out the probability distribution. You win \$16 or lose \$2.

x	$+16$	-2
$P(x)$	$\frac{1}{12}$	$\frac{11}{12}$

Calculate the expected value.

$$E(X) = 16 \times \tfrac{1}{12} + (-2) \times \tfrac{11}{12}$$
$$= -0.5$$

Write as an amount.

You would expect to lose 50c on each bet of \$2.

Express this as a percentage.

$$\text{House percentage} = \frac{0.50}{2.00}$$
$$= 25\%$$

b Choose a variable.

Let \$$a$ be the amount won on a bet.

Rewrite the probability distribution, remembering you get back \$$a$ + \$2.

x	$+a$	-2
$P(x)$	$\frac{1}{12}$	$\frac{11}{12}$

Recalculate the expected value.

$$E(X) = a \times \tfrac{1}{12} + -2 \times \tfrac{11}{12}$$

Write the value for a fair game.

To be fair, $E(X) = 0$

Solve the equation.

$$a \times \frac{1}{12} + (-2) \times \frac{11}{12} = 0$$
$$a = 22$$

Write the answer.

For the game to be fair, you should win \$22, so you should get back \$24.

Insurance companies use statistical information to work out premiums. Using expected values, they set the amount you pay for insurance to cover the expected payouts, costs and profits.

Example 19

In Australia, 831 males and 387 females aged between 15 and 24 died in 2011. These worked out to relative frequencies of 0.000 513 and 0.000 253 respectively.

a If there were no costs or profits, what should it cost to insure a male aged 23 for one year with a \$200 000 payout on death?

b An insurance company charges \$53 for \$100 000 cover for a 22-year-old non-smoking female. How much of this is to cover costs and profit?

Solution

a Choose the random variable.

Let X be the amount 'won' on death.

Let the premium be k. Write the probability distribution.

x	$200\,000 - k$	k
$P(x)$	0.000 513	0.999 487

Calculate the expected value.

$E(X) = 0.000\,513(200\,000 - k) + 0.999\,487k$

Write an equation for a fair amount.

$0.000\,513(200\,000 - k) + 0.999\,487k = 0$

Solve to find the premium.

$k \approx \$102.71$

Write the answer.

The premium would be about \$102.71.

b Choose the random variable.

Let X = amount paid by client

Write the probability distribution.

x	53	–99 947
$P(x)$	0.999 747	0.000 253

Calculate the expected value.

$E(X) = 0.999\,747 \times 53 + 0.000\,253 \times (-99\,947)$

Work out the amount.

$= 27.7$

Write the answer.

\$27.70 is expected for costs and profits.

INVESTIGATION Life insurance

Use information from the Australian Bureau of Statistics (Catalogue number 3303.0) to work out the probability of death for different age groups.

Use the internet or other methods to find the cost of life insurance for a year for different age groups and circumstances (gender, general health, life habits).

Compare the actual life insurance cost with the theoretical cost based on the probabilities of death. Why are there differences?

EXERCISE 2.08 Applications of discrete random variables

Reasoning and communication

1 Example 17 Three dice are rolled. It costs \$5 to play. If all 3 dice are the same, you get \$10 back. If 2 of the dice are the same, you get your \$5 back. Otherwise you lose.

a How much can you expect to win each time you play the game?

b Is the game fair?

2 Example 18 In a game of chance, a fair die is rolled three times. If three 4s are rolled, the player wins \$3. If the player rolls two 4s, \$2 is won and one 4 means the player wins \$1. If no 4s are rolled, then the player loses the \$1 bet. Calculate the house percentage for this game.

3 Example 19 An insurance company sells a term life insurance policy that will pay a beneficiary a certain sum of money upon the death of the policy holder. These particular policies have premiums that must be paid annually.
Suppose the company sells a \$300 000 one-year term life insurance policy to a 51-year-old male for an annual premium of \$650. According to actuarial tables, the probability that the 51-year-old male will survive the year is 0.99788. Calculate the expected value of this policy for the insurance company.

4 A local supermarket has four checkout lanes. The manager has determined that the number of lanes in use when it opens at 8:00 a.m. is a random variable having the following probability distribution.

x	0	1	2	3	4
$P(X = x)$	0.1	0.15	0.15	0.2	0.4

Calculate the expected number of checkout lanes in use at 9:00 a.m. at the supermarket.

5 New cars that come off an assembly line are fitted with five new tyres (including a spare). The probability distribution for the number of defective tyres on each car (D) is given by:

d	0	1	2	3	4	5
$P(D = d)$	0.95	0.03	0.015	0.003	0.0015	0.0005

a Find the expected number of defective tyres per car.

b What is the most likely number of defective tyres per car?

6 All of the households in a large regional town were surveyed to determine the number of persons living in the household. The results of the survey are shown below.

Number of persons	1	2	3	4	5	6	7
Number of households (1000s)	31.1	38.6	18.8	16.2	7.2	2.7	1.4

a What is the probability that fewer than 4 persons are in any given household?

b What is the expected number of persons in any given household?

c What is the standard deviation of the number of persons in any given household?

7 An individual plays a game in which it is possible to lose \$1, break even, win \$3 or win \$5. The probability distribution for the outcomes of this game are as follows:

Outcome (\$)	−1	0	3	5
Probability	0.3	0.4	0.2	0.1

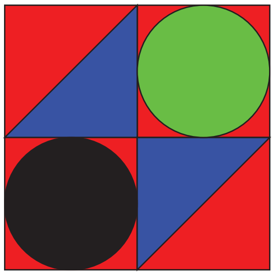
Does this game favour the player or the 'house'? Explain.

8 a The game described in the previous question is adjusted so that all payouts are reduced by \$1. How does this impact the expected outcome for the player?

b The adjusted game described in part **a** is further adjusted so that all payouts are doubled. How does this impact the expected outcome for the player?

9 A lottery is run to raise funds for a school building program. The lottery offers a grand prize of \$20 000, 20 prizes of \$500 each and tickets cost \$10. The lottery organisers know that 10 000 tickets will be sold. Determine whether purchasing a ticket is a good 'investment'.

10 Many games of chance are played in 'sideshow alley'. One game consists of throwing a dart that lands on a random spot within a square target as shown here. Each dart costs \$5. If the dart lands on the blue triangles, the player gets \$5 so comes out even. The black circle gets the player \$8, while the green circle gets the player \$10. Do you expect to make money or lose money when you play?



11 An insurance company sells term life insurance policies for a 20-year period. A 35 year-old woman takes out one of these policies and pays a premium of \$2500. If the insured person dies within the 20-year period, the insurance company pays out \$75 000. It is known that the probability that a 35-year-old woman will die within 20 years is 0.017. What is the amount that the insurance company can expect to gain from this policy?

12 An insurance company offers a policy that pays out \$50 000 on death. The company wishes to make a profit of \$225 on this type of policy. It is known that the probability of a male aged 38 dying is 0.009. What premium will the insurance company need to charge a 38-year-old male in order to achieve the desired profit?

13 Ms Hardsell, an insurance agent, offers a 1-year term life insurance policy to males in a particular age category. The cost of the insurance is \$30/thousand dollars of coverage. According to actuarial tables, the probability that a male in this category will die within the next year is 0.005.

a What is the expected gain for the insurer for each thousand dollars of coverage?

b If insurance is sold only in multiples of \$1000 and if the overheads for writing such a policy are \$70, what is the minimum amount of insurance that Ms Hardsell should sell in a policy in order to have a positive expected gain?

9780170250290

14 A worker employed to maintain quality control at an electronics plant is paid \$300 per day to test complex electronic equipment made on a production line. Only 5 units a day are produced and the quality controller is paid a bonus of \$200 for each defective unit found. The probability function for the number of defective units is given by:

x	0	1	2	3	4	5
$p(x)$	0.90	0.06	0.02	0.008	0.006	0.006

a Calculate the expected number of defective units per day.
b Find the quality controller's average daily salary.
c What is the quality controller's most likely daily salary?

15 A game involves throwing a pair of dice and adding the numbers on the top of the dice. You lose \$3 if you get a sum of 7 or 11, you win \$6 if both dice show the same number, and no money changes hands in any other case.
a Calculate your expected winnings if you play the game.
b If you had to pay an 'entrance fee' to play the game, what would you consider to be a 'fair' price?

16 In a particular business venture, a person can make a profit of \$10 000 with a probability of 0.2, a profit of \$5000 with a probability of 0.2 and a loss of \$1000 with a probability of 0.6. If X is the return that will result for the person, calculate:
a the expected return
b the most likely return.

17 Two teams, A and B, play in a tournament. The probability that A wins any game in the tournament is $\frac{1}{2}$. The first team to win two games in a row or a total of three games wins the tournament. Find the expected number of games in the tournament.

18 A class of 30 students sits for two separate maths exams. The results of the class for the first exam range from 31% to 96% while the results for the second exam range from 49% to 95%. Based on these results, what can be said about the standard deviation for the distribution of the results for the first exam compared with the standard deviation for the distribution of the results for the second exam?

19 A company is considering an investment and believes that the return in the following year will be a random amount X having the following probability distribution, where x is shown as thousands of dollars. Find the expected return.

x (\$1000s)	100	250	300	350	400
$P(X = x)$	0.15	0.35	0.25	0.15	0.1

2 CHAPTER SUMMARY
DISCRETE RANDOM VARIABLES

- A probability **experiment** has a number of **outcomes**. Each outcome is called a **sample point** and the **sample space** for an experiment consists of all possible outcomes. An **event** is a subset of the sample space.
- The **union** ($\cup$) of events A and B is the combination of either event A or event B or both A and B occurring. The **intersection** ($\cap$) of events A and B is the event that both A and B occur and includes the sample points that are common to A and B.
- The **probability** of event $A = P(A) = \dfrac{\text{number of favourable outcomes}}{\text{total number of possible outcomes}}$. For any event A, $0 \le P(A) \le 1$. If $P(A) = 1$, then event A is certain to occur. If $P(A) = 0$, then event A cannot occur. The sum of the probabilities of all events for an experiment $= \sum p(x) = 1$.
- The **complement** of event A is represented as A'. The probability of A' is defined as the probability that A will not occur so $P(A) + P(A') = 1$ or $P(A') = 1 - P(A)$.
- The **addition rule** of probability states that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- **Mutually exclusive** events cannot occur simultaneously, so for mutually exclusive events $P(A \cap B) = 0$ and $P(A \cap B) = P(A) + P(B)$.
- The **conditional probability** of event A following the occurrence of event B is written as $P(A \mid B)$ or 'the probability of A given B', and defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- The **multiplication rule** for any events A and B is $P(A \cap B) = P(A \mid B) \times P(B)$.
- For **independent** events $P(A \cap B) = P(A) \times P(B)$
- A variable whose value is determined by chance is said to be **random**. The values a variable can take is called the **domain** of that variable. A random variable is **discrete** if the domain is a finite or countable number of numerical values. A **continuous** random variable is one that can take any value based on measurement along a continuum.
- A **probability function**, $\boldsymbol{P(X = x)}$ or $\boldsymbol{p(x)}$, is a function of a discrete random variable that gives the probability that the outcome associated with that variable will occur. For any probability function:

$$0 \le P(X = x) \le 1 \text{ for all values of } x \text{ and } \sum P(X = x) = 1$$

- A **discrete probability distribution** shows a probability function typically as a table, list or graph.
- In a **uniform probability distribution**, all outcomes are equally likely, so:

$$P(X = x) = \frac{1}{n}, \text{ where } n = \text{the number of outcomes}$$

- For a uniform probability distribution of a discrete random variable, X, with n equally likely outcomes on the integers from a to b:

$$P(X = x) = \frac{1}{n}, \text{ where } n = b - a + 1$$

and

$$P(c \le x \le d) = \frac{d - c + 1}{b - a + 1} \text{ where } a \le c \le d \le b$$

- An integer **random number** is a number selected from a uniform discrete distribution on the integer limits of the number.

9780170250290

- A **hypergeometric experiment** is the sampling of a particular number of items from a finite population without replacement. The population is divided into two types, classified as success and failure, bad and good, or 1 and 0, with a fixed number of successes. The result of the experiment is the number of successes in the sample.
- The sample size of a hypergeometric experiment is usually written as n, the population size as N, and the number of successes in the population is often written as k or M.
- A **hypergeometric distribution** is the probability distribution of a hypergeometric experiment. The random variable is the number of successes. For the hypergeometric random variable X, $P(X = x) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$,

  where $x \leq n$ and $x \leq k$.
- The **expected value**, μ or $E(X)$, of the probability distribution of the discrete random variable X is the mean value of X: $\mu = E(X) = \sum xp(x)$
- The symbol $\sum$ means the sum of all values
- For the discrete random variable X the **variance, *Var*(*X*)**, and **standard deviation**, **σ**, are measures of the 'spread' or dispersion of X.

  $$Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$$
  $$\sigma = \sqrt{Var(X)}$$
- A game is fair if $E(X) = 0$, favourable to the player if $E(X) > 0$ and unfavourable to the player if $E(X) < 0$.

2 CHAPTER REVIEW
DISCRETE RANDOM VARIABLES

Multiple choice

1 Example 1 If $P(C) = 0.75$, $P(D) = 0.36$ and $P(C \cap D) = 0.29$, then $P(C \cup D)$ is equal to:

A 0.07 B 0.65 C 0.8 D 0.82 E 1.11

2 Example 1 If $P(M) = 0.56$, $P(N) = 0.24$ and $P(M \cap N) = 0.18$, find $P(M \mid N)$ correct to 2 decimal places.

A 0.11 B 0.32 C 0.43 D 0.47 E 0.75

3 Example 2 A fair coin is tossed three times and X = the number of heads. The probability function for X is:

A $P(1) = \frac{1}{8}, P(2) = \frac{3}{8}, P(3) = \frac{4}{8}$

B $P(1) = 0.2, P(2) = 0.3, P(3) = 0.5$

C $P(0) = \frac{1}{8}, P(1) = \frac{4}{8}, P(2) = \frac{2}{8}, P(3) = \frac{1}{8}$

D $P(0) = 0.1, P(1) = 0.2, P(2) = 0.3, P(3) = 0.4$

E $P(0) = \frac{1}{8}, P(1) = \frac{3}{8}, P(2) = \frac{3}{8}, P(3) = \frac{1}{8}$

4 Example 3 Which of the following is a valid probability value for a discrete random variable?

A 0.3 B 1.4 C -0.7 D $\frac{4}{3}$ E $\frac{11}{10}$

5 Example 3 A discrete variable can have:

A only integer values

B only specific values

C only rational values

D only integers or halves

E only non-negative integer values.

6 Example 4 Which one or more of the following could represent a probability distribution?

A

x	$p(x)$
0	0.4
1	0.3
2	0.2
3	0.1

B

x	$p(x)$
0	0.1
1	0.3
2	0.4
3	0.1

C

x	$p(x)$
0	−0.4
1	0.3
2	0.6
3	0.5

D

x	$p(x)$
0	0.2
1	0.2
2	0.2
3	0.2

E

x	$p(x)$
0	0.1
1	0.3
2	0.3
3	0.1

7 Example 9 A hypergeometric probability experiment with a population of 20 has a random variable, X. If the sample size is 4 and it is known that there are 8 successes in the population, what is the maximum value of x?

A 4 B 8 C 12 D 16 E 20

8 Example 10 A hypergeometric probability experiment is defined by $P(X=4)=\dfrac{\binom{10}{4}\binom{22}{8}}{\binom{32}{12}}$

What is the number of successes in the population?

A 4 B 8 C 10 D 22 E 32

Questions **9-10** refer to the following probability distribution.

x	1	2	3	4
$P(X=x)$	0.2	0.3	0.4	0.1

9 Example 13 $E(X)$ is equal to:

A 1.9 B 2.4 C 2.5 D 2.6 E 3.7

10 Example 14 $Var(X)$ is equal to:

A 0.78 B 0.84 C 1.04 D 1.22 E 1.46

11 Example 15 You have organised a camping weekend in the wet season. The Bureau of Meteorology says that there is a 50% chance of rain on the first day. If it rains on the first day, there is a 50% chance of rain on the second day. If it doesn't rain on the first day, there is only a 30% chance it will rain on the second day. How many rainy days can you expect on the weekend?

A 0.75 B 0.85 C 0.9 D 0.95 E 1.05

iStockphoto/LisaValder

Short answer

12 Example 1 Karla takes her car for a service every 6 months. The probability that the car will need a minor repair is 0.45 and the probability that the car will require a major repair is 0.1. The probability that the car will need both a minor and major repair is 0.04.

a What is the probability that the car will not require a minor repair but will need a major repair?

b What is the probability that the car will not need either a minor or major repair?

c Given that the car requires a minor repair, what is the probability that it will require a major repair?

13 Example 3 Classify each of the following variables as discrete or continuous.

a The number of people in a car

b The distance a car travels

c The number of trips by car in a week

d The time a car is out of the garage

e The number of cylinders in a car

f The dress sizes of women in a bus

g The weights of women in a bus

h The prices of dresses

14 Examples 2, 3 Which of the following could represent probability distributions?

a

x	$p(x)$
10	0.4
15	0.3
20	0.3

b

y	$p(y)$
−1	0.15
0	0.75
1	0.10

c

x	$p(x)$
1	0.7
2	−0.2
3	0.5

15 Example 4 Two probability distributions are defined by $p_1(x)=\frac{x}{10}$ and $p_2(x)=\frac{x^3}{6}+\frac{5x}{6}-\frac{x^4}{60}-\frac{7x^2}{12}-\frac{3}{20}$ for $x=1, 2, 3, 4$.

a Construct the probability distributions.

b Draw graphs of the distributions.

c Verify that each is a probability distribution.

d Is either distribution uniform?

16 Example 7 A coin is tossed 4 times. Let X be the random variable representing the largest number of successive heads that occur in 4 tosses.

a Construct the probability distribution for X.

b Draw a graph of the distribution.

17 Examples 7, 8 Five marbles in a bag are numbered:

2 7 8 10 12

D is the number of the marble randomly selected from the bag.

a Draw a graph of the probability distribution for D.

b Calculate the probability that the number on the ball that is drawn is odd.

c Describe the distribution.

9780170250290

18 Examples 6, 12 The numbers of tracks on some randomly selected CDs were as follows.

11	13	19	20	13	11	15	13	14	16	16	14	12
18	20	12	14	14	16	20	20	14	14	12	13	

a Use a frequency table to construct the probability distribution function for the number of tracks, T.

b What proportion of the CDs have exactly 12 tracks?

c What proportion have 15 or more tracks?

19 Example 9 For each of the following hypergeometric experiments, identify the random variable (X) and determine the values of N, n, k and the possible values of X.

a A production run of 40 components contains 7 defective ones. Five components are randomly selected and removed for testing.

b A manufacturer of car tyres reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly faulty. A tyre supplier purchases 10 of these tyres at random from the manufacturer.

c A workplace has 100 employees. It is known that 12% of the employees have blood type O. A group of 15 employees is selected from the workplace for a biological case study.

20 Example 10 A hypergeometric probability experiment is conducted with the given parameters below. Find the probability of obtaining x successes in each case.

a $N = 40, n = 10, k = 12, x = 5$ **b** $N = 65, n = 12, k = 25, x = 4$

c $N = 250, n = 60, k = 30, x = 15$

21 Examples 13,14 A pair of fair dice is rolled and the difference between the upper faces noted.

a Construct the probability function for D, the random variable of the difference.

b Find the expected value of D.

22 Examples 12–15 The results of a test out of 10 are shown below.

5	8	9	1	2	4	5	5	4	3	6
7	2	3	5	5	6	8	1	10	3	7
7	5	5	8	3	5	6	8	10	2	6

Find the expected value and variance of the variable T for the test results.

23 Example 15 Calculate $E(X)$ and the variance for the following probability distribution.

x	−2	0	2	4
$p(x)$	0.25	0.125	0.5	0.125

24 Example 15 The number of mobile phones sold by a dealership each month has the following probability distribution.

y	200	300	400	500	600	800
$P(Y = y)$	0.05	0.15	0.35	0.25	0.15	0.05

Calculate:

a $E(Y)$ **b** $Var(Y)$ **c** the standard deviation for Y.

Application

25 A coin is weighted so that $P(\text{H}) = \frac{3}{4}$ and $P(\text{T}) = \frac{1}{4}$. The coin is tossed 3 times and the results are recorded. If X is a random variable representing the greatest number of successive heads that occur,

a construct the distribution for X

b find the expected value of X.

26 Two normal dice are rolled and the values on the dice are multiplied to give a number between 1 and 36. Find the expected value and variance of the random variable. Is the distribution uniform or non-uniform?

27 In one day's production at a motor vehicle factory, 80 cars are produced. Of these, 18% require paintwork touch ups before they can be sent to dealers for sale. If a sample of 15 cars from the day's production is selected at random and tested, calculate the probability that:

a four will be defective

b at least three will be defective.

28 A particular game that operates from newsagencies involves scratching the special coating away from a panel to reveal 6 amounts of money. If 3 matching amounts are revealed, the player wins that amount.

In a certain series of the game, 10 million tickets are prepared and prizes allocated as follows:

1 @ \$100 000 25 000 @ \$100

2 @ \$25 000 25 000 @ \$10

25 @ \$1000 1 200 000 @ \$2

If it costs \$1 to play the game, calculate:

a the probability of winning a prize

b the expected payout on each ticket

c the house percentage.

123RF/Scott Rothstein

29 A dice game is played according to the following rules. Two dice are thrown. If the total is less than 6 or greater than 9, the player wins \$10. Otherwise, the player loses \$5.

a How much can you expect to win each time you play the game?

b Is the game fair?

c If the game is unfair, how could it be altered to make it fair?

30 In a particular business venture, a person can make a profit of \$10 000 with a probability of 0.2, a profit of \$5000 with a probability of 0.2 and a loss of \$1000 with a probability of 0.6. If X is the profit or loss that will result for the person, calculate:

a the expected value of X

b the most likely value of X.

31 In a game of chance, a player selects (without replacement) three balls from a bag that contains 3 white and 17 red balls. If a player selects 3, 2, or 1 white ball, then the player gets back \$5, \$2 and 50c respectively. If it costs the player 50c to play, how much do you expect the player to win or lose for each play?

32 A company manufactures a mobile phone case that it sells for \$6 each. Estimates of sales demand (the number of cases that will be sold) and variable costs (in dollars) for a sales period are determined to be as follows.

Sales demand	Probability
5000	0.3
6000	0.6
8000	0.1

Variable costs	Probability
\$3.00	0.1
\$3.50	0.3
\$4.00	0.5
\$4.50	0.1

In addition to the variable costs, there are fixed costs of \$8000 regardless of the number of units produced.

Calculate the expected profit for the sales period.

Practice quiz